ダイナミック・プログラミングとベイズ適応制御系 (動的計画法研究会報告集)

AUTHOR(S):

坂口, 実

CITATION:

坂口, 実. ダイナミック・プログラミングとベイズ適応制御系 (動的計画法研究会報告集). 数理解析研究所講究録 1967, 28: 61-73

ISSUE DATE:

1967-09

URL:

http://hdl.handle.net/2433/107532

数理解析研究所講究録
28巻 1967年 61-74

ダイナミック・プログラミングとベイズ適応制御系

阪大・基礎工　坂口　実

§1. 適応制御系

線型制御系

(1.1)　$x_n = \alpha x_{n-1} + v_n + \gamma_n \quad (n = 1, 2, \cdots;\ x_0 = c)$

を考える。ここに $\alpha \neq 0$ は与えられた定数，$\{\gamma_n\}$ は，独立確率変数列であって，同一分布 $N(\theta, \frac{1}{2})$ に従がい，平均値 θ の値は未知である。制御は

(1.2)　$v_n = v_n(x_{n-1}, \gamma^{n-1}), \quad (n = 2, 3, \cdots)$

すなわち $\gamma^{n-1} = \{\gamma_1, \cdots, \gamma_{n-1}\}$ を情報として利用し，θ を推定しながら制御する。

θ が未知だからベイズ的接近をとる。すなわち θ の事前分布を $N(\mu, \sigma^2)$ として，情報 γ^n を得た後の θ の事後分布は，やはり $N(\mu_n, \sigma_n^2)$，ただし

(1.3)　$\mu_n = t_n \bar{\gamma}_n + (1 - t_n)\mu, \quad \sigma_n^2 = \frac{1}{2n + \sigma^{-2}} \quad (n \geq 1)$

(1.4) $\qquad t_n = 2n(2n + \sigma^{-2})^{-1}, \quad \bar{r}_n = n^{-1}\sum_{j=1}^{n} r_j$

で与えられる。$n=1$ に対して v_1 をえらぶときには θ に関する情報は何もないわけであるが，仮定

(A.1) {はじめに pilot observation $r^\circ \sim N(\theta, \frac{1}{2})$ をとることが許されていて，第一段では r_1 の分布があたかも $N(r^\circ, \frac{1}{2})$ として既知のときのようにふるまうものとする。第2段以後の制御は(1.2)であって r^{n-1} には r° を含ませない。

を設けておく。

$V = \{v_n\}_{n=1}^{N}$ を(条件つき)制御政策という。制御の目的は

(1.5) $$\int E_\theta\Big[\sum_{n=1}^{N}(\lambda v_n^2 + a_n x_n^2)\,\Big|\, r^\circ, V\Big]\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\theta-\mu)^2}{2\sigma^2}}d\theta \longrightarrow \min_V$$

であるとする。ここに $\lambda, a_1, \cdots\cdots, a_N$ は与えられた非負定数であって，E_θ は r° を観察した条件のもとで，r^N についてとった期待値を表わす。

この問題に対し，(a)ベイズ制御政策，(b) $\lambda = 0$ のときの minimax 制御政策を求め，(c) $\{x_n\}_{n=1}^{N}$ が直接観察できなくて $y_n = x_n + \eta_n$，ただし $\eta_n \sim N(0, b^2)$ (b^2 は既知) は独立な random noise，が観察可能である場合について

も考える。

§2. ベイズ制御政策

ダイナミック・プログラミングの常とう的手段により

$f_{k,N}(c, \gamma^{k-1})$ …… 時刻 k において情報が $x_{k-1}=c$ および γ^{k-1} のとき、以後ベイズ制御政策（$N(\mu_{k-1}, \sigma^2_{k-1})$ についての）を用いるとき得られるべき期待値

とおくと，明らかに $k \geq 2$ に対して

$$f_{k,N}(c, \gamma^{k-1}) = \min_{v_k} E^{(k)}\left[\lambda v_k^2 + a_k(\alpha c + v_k + \gamma_k)^2 + f_{k+1,N}(\alpha c + v_k + \gamma_k, \gamma^k)\right],$$

ここに

$$E^{(k)}[\phi(\gamma_k)] \equiv \int \frac{1}{\sqrt{2\pi}\sigma_{k-1}} e^{-\frac{(\theta-\mu_{k-1})^2}{2\sigma^2_{k-1}}} d\theta \int \phi(\gamma_k) \frac{1}{\sqrt{\pi}} e^{-(\gamma_k-\theta)^2} d\gamma_k$$

$$= \int \phi(\gamma_k) d\gamma_k \cdot \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2_{k-1}+\frac{1}{2}}} e^{-\frac{(\gamma_k-\mu_{k-1})^2}{2(\sigma^2_{k-1}+\frac{1}{2})}}$$

を表わす。（A.1）により，もしも

（2.2） $\mu_0 = \gamma^0$, $\sigma_0^2 = 0$

と定義しておけば，（2.1）はすべての $k \geq 1$ に対して成立するわけである。

まず（2.1）を $k=N$ のときに解くと（$f_{N+1,N}(\cdot) \equiv 0$），

(2.3) $$v^{*}_{N,N}(x_{N-1}, y^{N-1}) = -\frac{a_N}{\lambda + a_N}\left(\alpha x_{N-1} + E^{(N)}(y_N)\right)$$

(2.4) $$f_{N,N}(x_{N-1}, y^{N-1}) = \frac{\lambda a_N}{\lambda + a_N}\left(\alpha x_{N-1} + E^{(N)}(y_N)\right)^2 + a_N V^{(N)}(y_N).$$

いま，一般に

(2.5) $$f_{k,N}(c, y^{k-1}) = A_k c^2 - 2B_k(y^{k-1})c + C_k(y^{k-1}),$$

$(k = 1, \cdots, N)$ とおくと，(2.1) よりつぎの諸式が得られる（もしも分母が0でなければ）．

(2.6) $$v^{*}_{k,N}(c, y^{k-1}) = \frac{-(\alpha c + E^{(k)}(y))(a_k + A_{k-1}) + E^{(k)}\{B_{k+1}(y^k)\}}{\lambda + a_k + A_{k+1}}$$

$$(k = 1, \cdots, N\ ;\ E^{(1)}(y) = \mu_0 = y^0)$$

(2.7) $$A_k = \frac{\alpha^2 \lambda (a_k + A_{k+1})}{\lambda + a_k + A_{k+1}}$$

(2.8) $$B_k(y^{k-1}) = \frac{\alpha\lambda}{\lambda + a_k + A_{k+1}} \cdot \left[-(a_k + A_{k+1})E^{(k)}(y) + E^{(k)}(y) + E^{(k)}\{B_{k+1}(y^k)\}\right]$$

$$C_k(y^{k-1}) = E^{(k)}\{C_{k+1}(y^k)\} + (a_k + A_{k+1})E^{(k)}(y^2) - 2E^{(k)}\{B_{k+1}(y^k) \cdot y_k\} - \frac{\left[(a_k + A_{k+1})E^{(k)}(y) - E^{(k)}\{B_{k+1}(y^k)\}\right]^2}{\lambda + a_k + A_{k+1}}.$$

(2.4)，(2.5) から

(2.7a) $$A_N = \frac{\alpha^2 \lambda a_N}{\lambda + a_N}$$

(2.8α) $\quad B_N(\gamma^{N-1}) = -\frac{\alpha\lambda a_N}{\lambda + a_N} E^{(N)}(\gamma),$

(2.9α) $\quad C_N(\gamma^{N-1}) = \frac{\lambda a_N}{\lambda + a_N}\left(E^{(N)}(\gamma)\right)^2 + a_N V^{(N)}(\gamma)$

であるから，数列 $\{A_k\}_1^N$, $\{B^{(k)}(\gamma^{k-1})\}_1^N$, $\{C_k(\gamma^{k-1})\}_1^N$ を定めることができる。かくて

[定理1] 想定 (A.1) のもとで，(1.1) に対する Bayes 適応制御政策 ($N(\mu, \sigma^2)$ についての) は，(2.6)～(2.9) で与えられる。

これはもちろん non-adaptive case (Sakaguchi, 1962) を特別な場合として含む。

(注意1) $\lambda = 0$, $a_1, \cdots\cdots, a_N > 0$ のときは Bayes 政策は one-step optimal である。何となれば，このとき (2.7)～(2.9α) より

(2.10) $$\begin{cases} A_k \equiv 0, \quad B_k(\gamma^{k-1}) \equiv 0 \quad (k = 1, \cdots, N) \\ C_k(\gamma^{k-1}) = \sum_{i=k}^{N} a_i V^{(i)}(\gamma_i) = \sum_{i=k}^{N} a_i \left(\sigma_{i-1}^2 + \frac{1}{2}\right) \end{cases}$$

となり，(2.6) より

(2.11) $\quad v^*_{k,N}(x_{k-1}, \gamma^{k-1}) = -\left(\alpha x_{k-1} + E^{(k)}(\gamma_k)\right)$

となって，N に無関係である。

さらにこのとき

(2.12) $\quad f_{1N}(x_0, \gamma^0) = A_1 x_0^2 - 2B_1(\gamma^0)x_0 + C_1(\gamma^0)$

$$= \frac{a_1}{2} + \sum_{i=2}^{N} a_i \left(\sigma_{i-1}^2 + \frac{1}{2}\right)$$

$$= \frac{a_1}{2} + \sum_{i=2}^{N} a_i \left(\frac{1}{2} + \frac{1}{2(i-1)+\sigma^2}\right)$$

となって，x_0, γ^0 および事前分布における μ に無関係であり，σ^2 にのみ依存するのである。

(注意2). $\lambda > 0$ のときの terminal control：

(2.13) $\qquad a_1 = a_2 = \cdots = a_{N-1} = 0, \quad a_N = 1$

を考えよう。このとき (2.7), (2.7α), (2.8), (2.8α) より

$$(2.14)\quad \begin{cases} A_k = \dfrac{\alpha^{2(N-k+1)}\lambda}{\lambda + 1 + \alpha^2 + \cdots + \alpha^{2(N-k)}} \\ B_k(\gamma^{k-1}) = \dfrac{-\alpha^{N-k+1}\lambda E^{(k)}(\gamma_k)}{\lambda + 1 + \alpha^2 + \cdots + \alpha^{2(N-k)}} \cdot \sum_{i=k}^{N} \alpha^{N-i} \end{cases}$$

さらに (2.6) より

$$(2.15)\quad v^*_{k,N}(c, \gamma^{k-1}) = \frac{-\alpha^{N-k}}{\lambda + \sum_{i=k}^{N} \alpha^{2(N-k)}} \left(E^{(k)}(\gamma_k) \sum_{i=k}^{N} \alpha^{N-i} + \alpha^{N-k+1} c\right)$$

がでる。何となれば，つぎの補題があるからである：

(補題1). $\qquad 1 \leq k \leq i \leq N$ に対して

$$E^{(k)} E^{(k+1)} \cdots E^{(i-1)} E^{(i)}(\gamma_i) = E^{(k)}(\gamma_k) = \mu_{k-1}$$

証明： $\qquad E^{(k)} E^{(k+1)}(\gamma_{k+1}) = \mu_{k-1}$

を示せばよい。(1.3) より

$$E^{(k)}E^{(k+1)}(\gamma_{k+1}) = E^{(k)}(\mu_k) = E^{(k)}\{t_k\bar{\gamma}_k + (1-t_k)\mu\}$$
$$= \frac{t_k}{k}\left((k-1)\bar{\gamma}_{k-1} + E^{(k)}(\gamma_k)\right) + (1-t_k)\mu$$
$$= \frac{t_k}{k}(k-1+t_{k-1})\bar{\gamma}_{k-1} + \left(\frac{t_k(1-t_{k-1})}{k} + 1 - t_k\right)\mu,$$

ところが (1.4) より

(2.16) $$t_k(k-1+t_{k-1})/k = t_{k-1}$$

であるから証明された。▨

(2.15) は N に依存するから terminal control の Bayes 政策は one-step optimal でない。しかも (2.11) と (2.15) とを比較すればわかるように，制御問題 (1.5) の Bayes 政策は λ につき ($\lambda=0$ において) 連続でない。

§3. $\lambda=0$ のときの minimax 制御政策

前節注意 1 で述べたように，$\lambda=0$ のときの Bayes 政策は (2.11) で与えられるが，いま

(3.1)
$$\begin{cases} \hat{v}_{1,N}(c, \gamma^0) = -(\alpha c + \gamma^0), \\ \hat{v}_{k,N}(c, \gamma^{k-1}) = -(\alpha c + \bar{\gamma}_{k-1}), \quad 2 \le k \le N \end{cases}$$

で定義される政策を

$$\hat{V} = \{\hat{v}_{k,N}(c, \gamma^{k-1})\}_{k=1}^{N}$$

とおく。つぎに想定（A.1）のもとで minimax 制御問題:

$$(3.2)\qquad \sup_{\theta} E_\theta\Big[\sum_{n=1}^{N} a_n x_n^2 \,\Big|\, \gamma^\circ, \nabla\Big] \longrightarrow \min_{\nabla}$$

に対して，$\hat{\nabla}$ が minimax なことを示そう。

(1.1) に (3.1) を代入すると $\hat{\nabla}$ による軌道は

$$\begin{aligned} x_n &= \alpha x_{n-1} + \hat{v}_n + \gamma_n \quad (n=1,2,\cdots;\ x_0=c) \\ &= \alpha x_{n-1} + \gamma_n - (\alpha x_{n-1} + \bar{\gamma}_{n-1}) \\ &= \gamma_n - \bar{\gamma}_{n-1} \qquad (\text{ただし } \bar{\gamma}_0 = \gamma^\circ \text{と定義する}) \end{aligned}$$

したがって θ に無関係に

$$(3.3)\qquad E_\theta\Big[\sum_{n=1}^{N} a_n x_n^2 \,\Big|\, \gamma^\circ, \hat{\nabla}\Big] = \frac{a_1}{2} + \sum_{n=2}^{N} a_n\Big(\frac{1}{2} + \frac{1}{2(n-1)}\Big)$$

が成立する。

一方 (1.1) に Bayes 制御 (2.11) を代入すると ∇^* による軌道は

$$\begin{aligned} x_n &= \alpha x_{n-1} + v_n^* + \gamma_n \\ &= \alpha x_{n-1} + \gamma_n - (\alpha x_{n-1} + \mu_{n-1}) \\ &= \gamma_n - \mu_{n-1} \qquad (\text{ただし } \mu_0 = \gamma^\circ \text{と定義にある}) \end{aligned}$$

したがって $n \geq 2$ に対して

$$E_\theta(x_n^2) = E_\theta\{(Y_n - t_{n-1}\bar{Y}_{n-1} - (1-t_{n-1})\mu)^2\}$$
$$= \frac{1}{2} + t_{n-1}^2 \frac{1}{2(n-1)} + (1-t_{n-1})^2(\theta-\mu)^2,$$

ゆえに

$$E_\theta\Big[\sum_1^N a_n x_n^2 \,\Big|\, Y^0, V^*\Big] = \frac{a_1}{2} + \sum_{n=2}^N a_n\Big\{\frac{1}{2}\Big(1+\frac{t_{n-1}^2}{n-1}\Big) + (1-t_{n-1})^2(\theta-\mu)^2\Big\},$$

したがって

$$(3.4)\quad \int E_\theta\Big[\sum_1^N a_n x_n^2 \,\Big|\, Y^0, V^*\Big] \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\theta-\mu)^2}{2\sigma^2}}\, d\theta$$
$$= \frac{a_1}{2} + \sum_2^N a_n\Big\{\frac{1}{2}\Big(1+\frac{t_{n-1}^2}{n-1}\Big) + (1-t_{n-1})^2\sigma^2\Big\}$$
$$= \frac{a_1}{2} + \sum_2^N a_n\Big(\frac{1}{2} + \frac{1}{2(n-1)+\sigma^{-2}}\Big)$$

となって，当然のことながらこれは (2.12) と一致する。

かくして

$$(3.5)\quad \lim_{\sigma^2\to\infty}(3.4) = \frac{a_1}{2} + \sum_2^N \frac{a_n}{2}\Big(1+\frac{1}{n-1}\Big)$$

を得る。(3.3) と (3.5) および よく知られた補題：

(補題2) decision rule δ^0 が minimax なための一つの充分條件は

$\exists\{\xi^{(n)}\}_{n=1}^\infty$: $\xi^{(n)}$ に対する Bayes decision rule $\delta^{(n)}$ がつぎを満足する。すなわち，すべての θ に対して

$$\overline{\lim_{n\to\infty}} \int r(\theta, \delta^{(n)})\, d\xi^{(n)} \geq r(\theta, \delta^0)$$

(ここに $r(\theta, \delta^0)$ は δ^0 の risk ft. である).

により, $\hat{\nabla}$ は minimax 制御政策であることがわかった。

さらに

(補題3) $\hat{\nabla}$ は admissible control policy である。

証明: 簡単のため control policy ∇ に対し

$E_\theta\left[\sum_1^N a_n x_n^2 \,|\, \gamma^0, \nabla\right] = R(\theta, \nabla)$ とかこう。また $N(\mu, \sigma^2)$ の分布を ξ とかこう。まず (3.3), (3.4) から

(3.6)
$$\begin{aligned} &\sigma \int \left(R(\theta, \nabla^*) - R(\theta, \hat{\nabla})\right) d\xi(\theta) \\ &= \sum_{n=2}^{N} a_n \sigma \left(\frac{1}{2(n-1)+\sigma^{-2}} - \frac{1}{2(n-1)}\right) \\ &= \sum_{n=2}^{N} a_n \frac{-\sigma^{-1}}{2(n-1)(2(n-1)+\sigma^{-2})} \xrightarrow[(\sigma\to\infty)]{} 0 \end{aligned}$$

である。いま $\hat{\nabla}$ かわりに admissible でないとすれば

$\exists \nabla'$: $R(\theta, \nabla') \leq R(\theta, \hat{\nabla})$, for all θ and $<$ holds for at least one θ.

$R(\theta, \nabla)$ の θ についての連続性から

$\exists (A, B)$, $\Delta > 0$; $R(\theta, \nabla') \leq R(\theta, \hat{\nabla}) - \Delta$, in $A \leq \theta \leq B$.

すると,

$$(3.7)\quad \sigma\int (R(\theta,\hat{\nabla})-R(\theta,\nabla'))\,d\xi(\theta) \geq \sigma\int_A^B$$

$$\geq \frac{\Delta}{\sqrt{2\pi}}\int_A^B e^{-\frac{(\theta-\mu)^2}{2\sigma^2}}d\theta \xrightarrow[(\sigma\to\infty)]{} \frac{\Delta}{\sqrt{2\pi}}(B-A)>0$$

ところで

$$\sigma\int\left(R(\theta,\nabla^*)-R(\theta,\nabla')\right)d\xi(\theta) = \sigma\int\left(R(\theta,\nabla^*)-R(\theta,\hat{\nabla})\right)d\xi$$

$$+\sigma\int\left(R(\theta,\hat{\nabla})-R(\theta,\nabla')\right)d\xi$$

の右辺は (3.6), (3.7) により充分大きな σ に対し正となる。これは ∇^* の Bayes policy なることに矛盾する。▨

以上まとめて

[定理2] 想定(A.1)のもとで minimax control problem (3.2) に対して，一つの minimax かつ admissible control policy は (3.1) で与えられる。

(注意) もしも (A.1) を想定しなければ，すなわち，pilot observation r^0 をとることが許されず，最初の r_1 の分布を $N(\theta,\frac{1}{2})$，$\theta\sim N(\mu,\sigma^2)$ とすると，Bayes policy ∇^+ はやはり (2.6)－(2.9) で与えられる。ただし，このとき (2.2) の代りに

$$(3.8)\qquad \mu_0=\mu,\quad \sigma_0^2=\sigma^2$$

と定義しておかねばならない。そしてこのとき (3.4) は

$$\int E_\theta\Big[\sum_1^N a_n x_n^2 \,\Big|\, r^0, v^+\Big] \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\theta-\mu)^2}{2\sigma^2}} d\theta$$

$$= a_1\left(\frac{1}{2} + \sigma^2\right) + \sum_{n=2}^{N} a_n \left(\frac{1}{2} + \frac{1}{2(n-1)+\sigma^{-2}}\right) \xrightarrow[(\sigma^2\to\infty)]{} \infty$$

となって収束しないのである。

§4. $\{x_n\}$ が直接 observe できない場合

control system (1.1) において，$\{x_n\}$ が直接 observe できなくて，その代りに

(4.1) $$y_n = x_n + \eta_n \qquad (n = 0, 1, 2, \cdots)$$

$\eta_n \sim N(0, b^2)$ indep., 同一分布，b^2 は既知が observable，つまり x_n は y_n を通して推定されるのみとすると，optimal control はどうなるか？

(4.1) を (1.1) に代入すると

$$y_n = \alpha y_{n-1} + v_n + r_n + \eta_n - \alpha\eta_{n-1},$$

y_0 is known.

となる。

$$\xi_n = r_n + \eta_n - \alpha\eta_{n-1}, \quad n = 1, 2, \cdots$$

とおくと，$\xi_n \sim N\left(0, \frac{1}{2} + (1+\alpha^2)b^2\right)$ で独立，同一分布．

$\{y_n\}$ が observable だから $\{\xi_n\}$ もそうである。ゆえに

$$v_n = v_n(y_{n-1}, \xi^{n-1}).$$

また

$$J[v_1, \cdots\cdots, v_N] \equiv E\Big[\sum_{i=1}^{N} (\lambda v_n^2 + a_n(y_n - \eta_n)^2) \,\Big|\, r^\circ, \nabla\Big]$$

$$= E\Big[\sum_{1}^{N} (\lambda v_n^2 + a_n y_n^2) + \sum_{1}^{N} a_n \{(-2x_n y_n - 2y_n^2) + y_n^2\} \,\Big|\, r^\circ, \nabla\Big]$$

$$= E\Big[\sum_{1}^{N} (\lambda v_n^2 + a_n y_n^2) \,\Big|\, r^\circ, \nabla\Big] - b^2 \sum_{1}^{N} a_n$$

(x_n と y_n とは独立だから) であるから §2 の議論はほとんど全部そのまま通用する。

(文 献)

M. Sakaguchi, Dynamic programming for the optimal control of linear stochastic sepotems, Mathematica Japonical, 7 (1961/62), 79-86.